

# 目录

<b>第一章 语言智能的概念及其发展</b> .....	<b>1</b>
1.1 语言智能的基本概念和相关术语的界定 .....	1
1.2 语言智能的研究内容 .....	5
1.2.1 人类的语言与智能研究 .....	5
1.2.2 语言智能技术研究 .....	6
1.2.3 语言智能的创新应用研究 .....	7
1.3 国内外语言智能的发展 .....	8
1.4 语言学和语言智能 .....	10
1.5 新文科背景下学习语言智能的重要意义 .....	11
思考与讨论 .....	13
参考文献 .....	13
<b>第二章 人类的语言与智能</b> .....	<b>16</b>
2.1 引言 .....	16
2.2 人类语言与动物交际 .....	18
2.2.1 (自然界中的)动物交际行为 .....	18
2.2.2 人类语言的特征 .....	20
2.3 语言与智能 .....	23
2.3.1 语言是人类智能的关键 .....	23
2.3.2 语言智能的获得 .....	24
2.4 语言智能的神经基础 .....	26
2.4.1 脑的结构 .....	26
2.4.2 语言网络 .....	29
2.4.3 语言与语言外网络 .....	32

2.5 处理自然语言的人工神经网络.....	33
思考与讨论.....	34
拓展阅读.....	34
参考文献.....	34
<b>第三章 语言的认知与计算.....</b>	<b>40</b>
3.1 语言与认知研究.....	40
3.1.1 认知科学.....	40
3.1.2 第一代认知科学.....	42
3.1.3 第二代认知科学.....	43
3.2 基于第二代认知科学的认知语言学.....	44
3.2.1 历史背景.....	44
3.2.2 语言的认知基础.....	45
3.2.3 语义和认知.....	48
3.2.4 认知语法.....	49
3.3 语言、认知与计算的融合.....	51
3.3.1 认知心理学中的整体语言观.....	51
3.3.2 认知心理学中的语言理解模型.....	54
3.3.3 语言认知机理与计算模型的融合.....	57
思考与讨论.....	58
参考文献.....	58
<b>第四章 语言数据和语言模型.....</b>	<b>63</b>
4.1 “语言数据”的概念.....	63
4.2 语料库：语言数据的仓库.....	64
4.3 语言数据的加工处理.....	66
4.3.1 语料预处理.....	67
4.3.2 语料标注.....	69
4.3.3 基准数据集和评测.....	74
4.4 语言模型.....	74
4.4.1 语言学的理论模型.....	75
4.4.2 统计学的概率模型.....	77

4.5	走向大数据的语言模型.....	79
4.5.1	独热编码和词嵌入.....	80
4.5.2	词袋模型和 Skip-Gram.....	81
4.5.3	神经语言模型.....	82
	思考与讨论.....	83
	参考文献.....	83
<b>第五章</b>	<b>知识表示与知识图谱.....</b>	<b>85</b>
5.1	知识表示.....	85
5.1.1	知识表示.....	86
5.1.2	知识表示的方法.....	88
5.2	知识图谱的基本内涵.....	100
5.2.1	知识图谱的定义.....	101
5.2.2	知识图谱的类型.....	103
5.3	知识图谱的发展历史.....	106
5.4	知识图谱的主要技术.....	108
5.4.1	知识建模.....	109
5.4.2	知识获取.....	110
5.4.3	知识融合.....	114
5.4.4	知识存储与知识检索.....	114
5.4.5	知识推理.....	117
5.5	知识图谱的应用.....	117
5.6	知识图谱资源.....	119
	思考与讨论.....	121
	参考文献.....	121
<b>第六章</b>	<b>让机器识别和生成语音.....</b>	<b>123</b>
6.1	语音信号的声学基础及模型.....	123
6.1.1	语音信号的产生.....	123
6.1.2	语音听觉系统.....	125
6.1.3	语音信号的数学模型.....	126
6.1.4	语音信号的时域、频域特性及语谱图.....	127

6.1.5	汉语中语音的分类 .....	131
6.2	语音识别技术基础理论 .....	133
6.2.1	语音信号的预处理 .....	134
6.2.2	语音信号的端点检测 .....	136
6.2.3	语音识别特征提取 .....	137
6.3	传统语音识别 .....	140
6.4	端到端语音识别 .....	144
6.5	语音合成 .....	148
6.5.1	语音合成的基本原理 .....	148
6.5.2	典型的语音合成器模型 .....	150
	思考与讨论 .....	157
	参考文献 .....	158
<b>第七章</b>	<b>让机器理解和生成文本 .....</b>	<b>161</b>
7.1	自然语言处理 .....	161
7.2	自然语言处理的基本过程 .....	163
7.2.1	数据获取 .....	163
7.2.2	基础任务 .....	164
7.2.3	应用任务 .....	165
7.2.4	应用领域 .....	166
7.3	文本处理的主要任务 .....	167
7.3.1	命名实体识别 .....	167
7.3.2	情感分析 .....	171
7.3.3	文本分类 .....	175
7.3.4	自动文摘 .....	177
7.3.5	自动生成新闻 .....	180
	思考与讨论 .....	184
	参考文献 .....	184
<b>第八章</b>	<b>让机器与人进行言语交流 .....</b>	<b>187</b>
8.1	人机对话系统的发展历史和现状 .....	187
8.1.1	人机对话系统的分类 .....	187

8.1.2	人机对话系统的发展历史和现状 .....	189
8.2	相关理论与技术 .....	192
8.2.1	对话管理 .....	192
8.2.2	自然语言生成 .....	193
8.3	传统的人机对话系统 .....	195
8.3.1	问答系统 .....	195
8.3.2	对话系统 .....	200
8.3.3	闲聊系统 .....	207
8.4	基于大语言模型的人机对话系统 .....	213
8.4.1	基于大模型技术的人机对话系统的优势与不足 .....	216
8.4.2	提高大语言模型问答能力的方法和技术 .....	217
	思考与讨论 .....	228
	参考文献 .....	229
<b>第九章</b>	<b>让机器实现多语言翻译 .....</b>	<b>232</b>
9.1	机器翻译的概念、背景和历史发展 .....	232
9.2	机器翻译的原理与技术 .....	235
9.2.1	基于规则的机器翻译 .....	235
9.2.2	基于语料库的机器翻译 .....	237
9.2.3	基于深度学习的机器翻译 .....	241
9.2.4	混合机器翻译 .....	250
9.3	机器翻译质量测评 .....	251
9.3.1	自动化评价方法 .....	252
9.3.2	半自动测评方法 .....	255
9.4	低资源语言机器翻译 .....	256
9.4.1	低资源语言机器翻译方法 .....	257
9.4.2	低资源语言机器翻译实例 .....	262
9.5	经典机器翻译工具 .....	264
9.5.1	谷歌神经机器翻译 .....	264
9.5.2	DeepL .....	266
9.5.3	百度翻译 .....	267

9.5.4 小牛翻译 .....	267
思考与讨论 .....	268
参考文献 .....	268
<b>第十章 语言智能的新发展：大语言模型技术 .....</b>	<b>275</b>
10.1 大语言模型的发展历程和工作原理 .....	275
10.1.1 大语言模型的基本概念 .....	275
10.1.2 大语言模型的发展历程 .....	277
10.1.3 大语言模型的工作原理 .....	278
10.1.4 大语言模型的技术基础分析——以 InstructGPT/ ChatGPT 为例 .....	279
10.2 大语言模型的未来研究方向 .....	285
10.2.1 大模型“涌现”能力的原理与价值观对齐研究 .....	285
10.2.2 大语言模型的语言与评测研究 .....	286
10.2.3 数据标注与提示语工程的研究 .....	287
10.3 大语言模型文本生成功能的应用 .....	290
10.3.1 事件抽取 .....	290
10.3.2 文本分类 .....	293
10.3.3 受控文本生成 .....	299
10.3.4 意图识别 .....	305
思考与讨论 .....	307
参考文献 .....	307
<b>第十一章 语言智能伦理与规范 .....</b>	<b>310</b>
11.1 语言智能与伦理 .....	310
11.1.1 伦理的相关概念 .....	310
11.1.2 语言智能伦理 .....	311
11.2 隐私伦理问题 .....	312
11.2.1 隐私及隐私权 .....	312
11.2.2 隐私泄露 .....	312
11.2.3 隐私伦理问题的应对 .....	313

11.3	算法伦理问题 .....	314
11.3.1	偏见、歧视 .....	314
11.3.2	算法伦理问题的应对 .....	315
11.4	技术伦理问题 .....	316
11.5	学术伦理问题 .....	317
11.6	语言智能伦理规范准则 .....	318
11.6.1	国内语言智能伦理规范准则 .....	318
11.6.2	国际语言智能伦理规范准则 .....	319
11.7	规范准则对语言智能研究和应用的影响 .....	320
11.8	语言智能伦理问题防范治理的未来展望 .....	321
	思考与讨论 .....	322
	参考文献 .....	322



# 第一章 语言智能的概念及其发展

## 本章提要

语言智能作为智能科学与技术的重要组成部分，代表着人工智能发展的前沿水平。近年来，大语言模型的出现，使语言智能的发展突飞猛进，深刻影响着人类社会和人类文明，也使语言智能这一领域备受关注。本章首先对语言智能的概念和内容进行阐述，勾勒出语言智能的基本理论框架和研究内容。语言智能研究包括人类的语言与智能、机器的语言智能处理技术以及这些技术的创新应用。语言、认知和计算是语言智能的三大要素。然后简单概述语言智能在国内外的的发展情况，最后论述了新文科建设背景下学习语言智能的重要意义。时代呼唤新型交叉学科解决现实问题，语言智能学科的设立和发展，对我国人工智能发展战略实施和高端复合型人才培养具有重要的推动作用。

## 1.1 语言智能的基本概念和相关术语的界定

近年来，随着以 ChatGPT 为代表的大语言模型的出现，语言智能成为人们关注的焦点和热点，成为异常火爆的词汇。语言智能是一个多层次多领域的概念，涉及语言学、认知科学、计算机科学、人工智能等领域。关于语言智能的概念，不同学者从不同的角度进行过界定。霍华德·加德纳（Howard Gardner, 1983）在他的多元智能理论中曾使用“语言智能”这一术语，认为语言智能是灵活掌握语音、语义、语法，具备用言语思维、用言语表达思想和交际意图的能力，是人类多元智能的一个方面。随着计算机科学技术和人工智能的发展，人们一直在探索如何让机器也具有

像人类一样的语言能力，机器翻译、自然语言处理、人机对话等都是在朝着这一目标努力，让机器具有语言智能。在此背景下，语言智能这一术语越来越多地被用来特指计算机智能处理人类语言的能力和技术。周建设等（2017；2018；2019）认为，语言智能即语言信息的智能化，指运用计算机信息技术模仿人类的智能，分析和处理人类语言的过程。语言智能旨在运用计算机技术和信息技术，让机器理解、处理和分析人类语言，实现人机语言交互（胡开宝，田绪军，2020：59），使得机器在一定程度上拥有理解、应用和分析人类语言的能力（胡开宝，尚文博，2022：104）。

在计算机科学与人工智能领域，语言智能处理主要体现为自然语言处理，是指利用计算机等工具分析和生成自然语言（包括文本、语音等），从而让计算机“理解”和“运用”自然语言，可以让人类通过自然语言的形式与计算机系统智能交互（黄河燕，史树敏等，2020）。

“语言智能”在语言教育和语言学习领域也经常提及和使用。周建设（2019：45）指出，“语言智能有力促进了语言教学、语言学习的智能化，拓展了语言学研究的新领域，在未来教育中将发挥越来越重要的作用。”黄立波（2022：5）认为，语言智能研究旨在借助现代技术，尤其是人工智能技术，模拟人类的语言能力，实现语言学习的全程智能化。

可以看出，大部分学者是从机器智能的角度来界定“语言智能”概念，把它看成是与感知智能、行为智能、类脑智能并列的一种机器智能，即强调人工智能技术在语言相关领域的应用。梁晓波、邓祯（2021：85）认为，语言智能指的是“运用计算机信息技术模仿人类的智能，以及分析和处理人类语言文字、声音、知识、情感的科学技术”。周建设（2019：45）认为，语言智能研究是语言学与人工智能的交叉学科。简单而言，语言智能就是语言加人工智能，语言智能研究就是人工智能技术应用于自然语言处理领域的相关研究（黄立波，2022：4）。

我们认为，“语言智能”是智能科学的重要组成部分。智能科学研究智能的基本理论和实现技术，是由脑科学、认知科学和人工智能等学科构成的交叉研究领域（史忠植，2013）。因此，语言智能的研究也至少应在脑科学、认知科学和人工智能技术三个层面上进行，既有机理和模型的探讨，也有功能的仿真，才能形成完备的理论和应用互益发展的体系。

脑科学从大脑的结构和功能的角度揭示语言习得和使用过程中的神经活动规律，是语言智能研究的基础。认知科学研究人类智能的本质和规律，探讨语言使用过程中的概念语义形成的机制和过程、具身性特征等。人工智能通过语音识别、自然语言处理，知识图谱等技术来模拟和实现人类的语言能力，三个方面互相支撑，彼此促进，共同构成语言智能的主要学科基础。

从智能行为的主体来看，主要包括两类，以人类为代表的生物和以计算机为代表的机器，因此智能有两种表现形态，人类智能（Human Intelligence）和人工智能（Artificial Intelligence），后者以前者为基础和目标，是对前者的延伸和拓展。无论是人类智能还是人工智能，都涉及语言的使用和处理问题。在人类智能中，知识总是用自然语言（其中主要是文字）表示的，推理是人脑的功能，所以人类智能的标志是自然语言知识+人脑推理（李冠宇，黄映辉，2009）。人类主要以语言媒介来获取知识，对世界形成概念和认知。在知识应用阶段依然是用语言来表达思想和概念，语言、认知和思维密不可分。语言的学习和使用是人类的智能活动。语言生成和理解的研究需要借助脑科学、神经认知科学等学科领域的研究。在机器智能中，机器要与人进行交互，也要理解和生成自然语言，因此自然语言处理被称为人工智能皇冠上的明珠。语言智能研究人的语言能力和机器对人的语言能力的模拟和延伸，以及二者的互动同构关系。

语言智能不是单纯的机器智能，而是以人类语言能力为基础、实现增强与互联的深度人机结合的综合智能行为。对人类语言机制的深入探讨将推动机器认知智能的进一步发展。语言智能是机器智能走向认知智能的必经之路，自然语言处理是机器智能的关键核心技术，是制约人工智能发展的主要瓶颈。语言智能是人工智能的重要组成部分及人机交互认知的重要基础和手段。语言智能基于人脑生理属性、言语认知路径、语义生成规律，利用大数据与人工智能技术，构拟人机语义同构关系，让机器实施类人言语行为（唐良元，2023）。另外，语言智能还有力促进语言教学、语言学习的智能化，拓展语言学研究的领域，在外语教育和人才培养过程中发挥越来越重要的作用。

在理解语言智能概念的时候，我们通常会联想到与它相关的一些术语，例如，计算语言学、自然语言处理、语言技术、语言数据科学、语言资源学等。

计算语言学（Computational Linguistics）通常被认为是语言学或应用语言学的一个分支领域。它从计算的视角，研究自然语言的规律，提供各种语言现象的计算模型，最终目标是让计算机能够像人类一样分析、理解和处理自然语言。与计算语言学最相关、近乎可以替换使用的术语是“自然语言处理”。自然语言处理（Natural Language Processing, NLP）研究人与计算机之间用人类语言进行沟通交流的实现方法。虽然这一领域的研究涉及自然语言，即人们日常使用的语言，它与语言学的研究有着密切的联系，但它的重点在于研制实现自然语言理解和生成的计算机系统，特别是其中的软件系统。因而它通常被认为是计算机科学的一部分。计算语言学和自然语言处理研究的内容大部分重叠，但侧重点各不相同，它们是语言智能的重要实现途径。

语言技术（Language Technology）是指利用自然语言处理、机器学习和深度学习等技术，使计算机能够理解、解析和生成自然语言的能力。通过这些技术，计算机可以分析文本、感知语音以及与人类进行交流，核心目标是建立智能对话系统，使机器具备像人类一样的沟通能力。

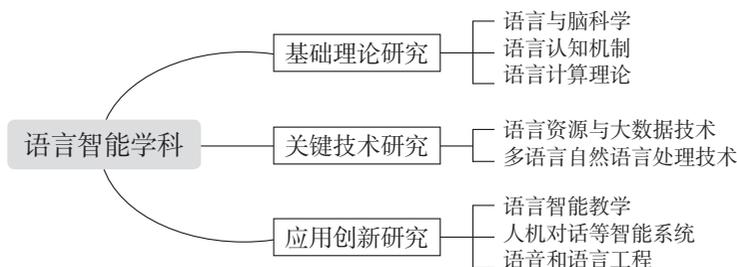
语言数据科学（Linguistic Data Science）是数据科学的一个领域。数据科学使用科学的方法，计算过程、算法和操作系统从结构化和非结构化数据中获得知识和观点。语言数据是指以语言为主要信息载体的数据，包含语言的各种属性，如词性、语法、句式、语义、语用等，形式可表现为文本数据、语音数据、视频数据等形式。在人工智能领域，语言数据常用于训练自然语言处理模型，如机器翻译、语音识别、文本分析等。

语言资源学（Study of Language Resources）是研究语言资源及其相关问题的科学，以语言资源为研究对象，系统研究语言资源的类型、构成、分布、质量特征、使用状况及其与语言研究和社会发展之间的关系，内容主要涉及语言资源理论、语言资源应用技术、语言资源管理与伦理等，是跨语言学、资源学、信息科学等的交叉学科。

这些术语和领域都和语言智能密切相关，并且在发展过程中深度融合，在推动语言智能发展的过程中发挥着重要作用。

## 1.2 语言智能的研究内容

从上一节可以看出，语言智能研究人的语言能力和机器对人的语言能力的模拟和延伸，以及二者的互动同构关系。它基于人脑生理属性、具身的认知体验、言语生成规律，利用大数据与人工智能技术，对自然语言信息进行处理，让机器实施类人言语行为，服务于人类社会。据此，语言智能的研究内容可以分为三个层次：基础理论研究、关键技术研究 and 应用创新研究。这三个层次主要内容之间的逻辑关系主要表现为：首先，语言的脑神经认知机制为语言智能提供关于人类语言处理的（生理、心理、语言）理论基础；其次，语言数据技术和计算模型为语言智能提供关于自然语言处理的关键技术支撑；第三，语言智能教学和人机互动系统等是语言智能技术的重要应用领域。这三大方面的研究密切联系，有机融合，共同构成了语言智能的整体框架。如下图所示：



### 1.2.1 人类的语言与智能研究

语言智能研究不仅涉及运用人工智能技术对语言信息进行智能化处理，而且要致力于研究人类大脑处理自然语言的神经认知机制，致力于探索语言的脑认知奥秘，以语言认知、语言学习、语言模型研究为重点，努力推动服务于认知智能、类脑智能发展的基础理论研究。苏金智（2021a）指出，语言智能与语言能力的关系很难得分一清二楚。他认为，人类的语言智能实际上就是人类语言能力的体现，机器人的语言智能就是机器人语言能力的体现，语言能力在这里起着关键的作用。

语言智能的基础理论研究主要包括以下内容：

(1) 语言与脑科学研究，主要关注人类语言处理的神经生理基础，研究大脑语言功能发展、多语言学习的脑加工机制、语言病理机制，以及语言产生、接收、分析和储存的神经机制等。

(2) 语言认知机制研究，主要关注人类语言处理的认知心理基础，研究语言理解、生成与使用的认知加工机制，语言能力和言语交际的认知解释等。

(3) 语言计算理论研究，主要关注人类语言的本质和系统特征以及面向计算机智能处理的理论和分析方法、构建语言模型等。语言理解是语言智能的主要任务，涉及语言理解的目标、途径和主要模型。机器对语言的理解就是确定了概念与语言单元的映射，以及知道承载概念单元的语言属性和知晓不同语言单元之间的关系。而语义理解是语言理解的核心，它包括对自然语言知识和常识的学习；语义理解可以通过一系列的人工智能算法以及多场景智能语义分析，将文本解析为结构化的、机器可读的意图与词槽信息。

### 1.2.2 语言智能技术研究

语言智能技术主要是计算机领域或人工智能领域感兴趣的话题，包括语言资源与大数据技术、多语言自然语言处理技术等。语言智能关键技术研究的主要目标是构建数据驱动、以自然语言理解和生成为核心的计算模型及优化方案，为语言计算、语言教育和语言工程应用提供关键技术支撑。

(1) 语言资源与大数据技术：利用计算机对语言数据进行抽取、存储、标注与研究，智能高效地处理语言大数据，构建语言资源。李宇明（2022）指出，“语言资源建设的科学目标，是满足计算机发展语言智能、从事各种机器语言行为的需要，满足计算机‘社会计算’的数据需要”。实现语言数据与语言知识、世界知识的有机融合，是语言智能处理的发展方向。语言数据是语言智能的基础，高质量的语言数据和资源是决定语言智能水平的重要因素。同时它也是语言研究的重要资源和基础。一方面，语言智能需要采用文本挖掘、知识图谱、深度学习、多模态识别等人工智能技术构建多语言语料库、知识库、资源库，帮助解决现实中的语言问题；

另一方面,利用语言数据挖掘和知识抽取算法发现新的语言现象,提出新的语言问题,从中总结出新的语言规律。

(2) 自然语言处理技术:以计算机为基本工具,研究自然语言的分析、理解和生成技术,以及面向语言的机器学习方法,实现有效的自然语言语义和意图计算。开展多语种、跨语言、多模态的智能语言处理实践,研究实现人机之间用自然语言进行有效沟通的各种方法和技术,致力于自然语篇语义理解、语言认知机制的计算建模等前沿研究,探索提升话语计算、情感分析、机器翻译准确率和精确度的有效路径,这些都是自然语言处理需要解决的问题。在大语言模型盛行的当今时代,探索大模型的原理和有效利用是这一领域关注的重点。

(3) 人机交互技术(Human-Computer Interaction, HCI):通过软硬件设计和实现来支持人与计算机之间交互的学科、过程和方法。它旨在促进用户友好的界面和系统,使人们能够尽可能轻松地使用计算机。该领域关注如何设计、评估、部署和使用计算机系统,以便人类用户可以更好地完成任务,并从中获得良好的用户体验。

### 1.2.3 语言智能的创新应用研究

语言智能的应用范围相当广泛,主要包括语言智能教学、人机对话等智能系统、语音工程与语言工程等,用于解决各领域与语言相关的实际问题。创新语言智能技术的应用研究、促进语言教育智能化、提升各类智能系统和平台的性能和质量、开发和拓展各类语言工程应用,也是语言智能学科需要研究的重要内容。

(1) 语言智能教学:开展人机共生时代语言教学理论、学习理论与学习者研究,创新智能化语言教学模式和方法,设计、管理和优化语言学习过程,开发个性化的多语言智能教学系统、测评系统和学习平台,建设数字化语言学习资源库,培养掌握智能技术和语言技能的复合型人才,提升语言教师的智能化教学素养等。

(2) 人机对话等智能系统:人机对话系统是语言智能的重要应用领域,近年来的最新进展主要来自深度学习技术的成熟和大规模对话语料获取这两方面的进步,但目前仍存在诸多问题和挑战(黄河燕,史树敏等,2020:196-197):在系统层,需要构建统一的人机对话框架;在模型

层,构建完全端到端的人机对话系统仍面临巨大挑战;在学习层,需要超越当前的有监督学习范式;此外,还需要与常识、现实世界和情境建立联系。

(3) 语音与语言工程:语音工程是与语音智能处理有关的实际应用,在语音产生、传递和感知研究的基础上,利用人工智能技术进行语音识别、语音编码、语音合成、言语感知等。语言工程主要应用语言智能解决各类现实问题。

随着语言智能技术的不断进步,应用场景会不断拓宽。在智能客服,多语言服务、数字人文、社会计算和国际传播等领域具有广阔的应用空间。加强语言智能的领域应用研究,与技术研发同等重要。

### 1.3 国内外语言智能的发展

语言智能研究在整个人工智能领域发挥关键核心作用。在国外,和语言智能密切相关的研究很早就受到了关注和重视,具备了较好的研究基础,但没有使用“语言智能”这一名称,较多地采用“语言技术”“语言与语音处理”“计算语言学与信息处理”等名称。国外的语言智能研究主要体现为语言技术研究与自然语言信息处理。例如,美国卡内基梅隆大学计算机科学学院设有语言技术研究所;斯坦福大学人工智能实验室设有专门的NLP研究组;麻省理工学院设有计算机科学与人工智能实验室,研究方向包括人机互动、机器人学等;约翰斯·霍普金斯大学设有语言与语音处理中心CLSP;马里兰大学设有计算语言学与信息处理实验室;宾夕法尼亚大学设有自然语言处理研究组(UPenn NLP Group);英国爱丁堡大学设有语言、认知和计算研究所。事实上,美国军事领域很早便启动了语言智能处理项目,整合了统计学、计算机科学、认知科学、语言学等领域的先进成果,目前在语言智能处理领域处于全球领先地位。

在我国,“语言智能”这个概念是在2013年北京“语言智能学术论坛”上提出的(周建设,2023)。随着深度学习、语言大数据技术的不断发展,越来越受到学界业界的重视并取得了令人瞩目的成就。在语言资源建设、多语言机器翻译、多语言文本挖掘、智能语言教学、智能舆情监控等领域发展迅速,为全球语言服务发挥了重要作用。

在理论建构和学科建设方面,学者们围绕语言智能的概念与内涵、语言智能与大数据、语言智能与国家语言能力、语言智能和语言学的关系、语言智能在外语教学中的应用、语言智能的学科建设与人才培养等进行探讨,使语言智能的概念逐渐广为接受,其重要意义受到前所未有的关注。语言智能是人工智能走向认知智能的关键所在,对于我国的人工智能发展战略至关重要,也是语言类学科参与重大科研攻关和社会服务、更新人才培养理念的重要契机。

目前,我国的语言智能学科建设已经具备了较好的基础,形成了较明确的研究方向,具备较为稳定的教学、科研队伍,也产出了很多技术应用成果。在我国的学科体系中,语言信息处理一般是语言学及应用语言学、计算机应用技术、教育技术学等二级学科下的研究方向。近年来,国内一些高校新设置了语言智能相关学科或研究机构,分属中国语言文学、计算机科学与技术或外国语言文学等一级学科,如首都师范大学、北京语言大学、重庆大学、四川外国语大学等。上海外国语大学、北京外国语大学、青岛大学、北京理工大学、西安外国语大学等高校也在积极开展语言数据科学、“人工智能+外语”等新文科探索和语言智能相关研究。这些进展充分说明,语言智能交叉学科建设在我国已受到较广泛的重视,事实上已经成为新文科建设的一个重要方向。

从技术演进的角度来看,语言智能(机器翻译、自然语言处理、人机对话、语言识别等)的发展脉络和语言模型的范式演进过程趋同。语言模型的“范式”通常指的是模型的基本建构方式。范式决定了语言模型如何捕获语言单位的概率分布和语法结构,以便执行不同的自然语言处理任务。从自然语言处理的研究范式来看,1950—1990年占主导地位的是小规模专家知识,1990—2010年是浅层机器学习,2010—2017是深度学习,2019—2022是预训练语言模型,2023年开始进入大模型时代(刘挺,2023)。在这过程中逐渐形成的模型范式主要包括基于规则的范式、基于概率统计的范式、基于深度学习的范式和预训练+微调范式,语言模型的范式演进过程贯穿了语言智能的发展过程。具体内容因为后面章节都有所涉及,此处暂不展开讲解。

## 1.4 语言学和语言智能

讨论人类的语言智能也好，机器的语言智能也好，都必然会涉及语言的原理和规律问题，只有深刻认识人类语言的本质属性和使用的机制，才能更好地深入研究语言智能的前沿领域。语言学作为一个蓬勃发展的学科，主要研究和描写人类语言的结构、功能及其历史发展，揭示语言的本质，探索语言的共同规律。语言是人类所独有的，只有人类才使用有声语言进行交际，因此，通过语言的研究，我们可以更加清楚透彻地理解人类的本质。语言是人类进化过程中的重要成果之一，而语言学正是将人类语言从古至今的演化历程进行系统、全面、深入地研究，并得出有意义的结论，从而让我们更好地了解人类智慧和文明的发展历程。比如，语言学家们通过对各种语言结构的形式和规律的探讨，可以揭示不同文化背景下人们的思维方式和行为模式，为世界各地的跨文化交流提供理论基础。

语言学与语言智能有密切关联，并在语言智能的发展过程中发挥重要作用。机器智能的探索始于对机器翻译的研究，在最初基于规则的机器翻译发展阶段，语言学知识和理论发挥着主导作用。自然语言处理、人机对话等领域也大都采用语言学家编制规则的方法来实现智能操作。随着深度学习的应用、计算机软硬件的不断发展，语言智能愈来愈依赖于语言数据，而非语言学知识，但效果越来越好，不仅能够生成符合语法的句子，还可以更好地“理解”人类语言。基于大数据和深度学习的自然语言处理技术已成为语言智能的主流。语言学对于语言智能的重要性似乎逐渐下降，甚至有人认为语言学研究对于语言智能发展的意义并不大。

语言智能是由多学科构成的交叉领域，其中计算机科学与技术、认知科学等发挥关键作用，但语言学的作用不可或缺。没有语言学的理论和研究成果作指导就不可能有语言智能技术的创新和发展。语言学对语言结构的描写和语言使用规律的总结可以应用于语言数据的标注。语义学的研究可应用于知识图谱的建构的设计。言语交际的理论可应用于人机对话的改进和提升，认知语言学的理论可以指导人工智能走向具身认

知智能等等。事实上，无论是过去还是将来，语言智能的发展离不开语言学的支撑，人工智能算法并不能代替语言学理论（胡开宝，2023）。

当然，语言智能发展到今天，特别是大语言模型的快速发展，确实也给语言学的研究提出了新的问题和挑战。例如，以 ChatGPT 为代表的大语言模型到底学到了哪些语言知识？为什么在不需要大脑思考的情况下，仅依靠大数据和大算力就可以理解和生成话语？我们应该如何刻画和评测大语言模型所表现出来的语言能力？在过去的语言学研究过程中，语言学家大多依赖语言直觉或有限的语言数据作为分析和总结规律的依据；在互联网高度发达且语言数据迭代增强的当今时代，能否借助人工智能和大数据分析技术，来重新建构我们对人类语言的观念和认知？在海量数据和人工智能加持下得出的语言学理论和规律会和我们以前的认知假设一致吗？语言学理论在语言智能发展的过程中一直发挥重要的支撑作用，反过来，语言智能的发展也会极大地促进语言学的研究和迭代升级，二者之间相互借鉴，相互促动，共同推动语言科学和技术的不断发展。

## 1.5 新文科背景下学习语言智能的重要意义

随着智能科学与技术的日益普及和广泛应用，社会对既通语言又懂技术的语言智能人才需求量越来越大。刘利（2019）认为，“当前语言智能领域对语言知识和语言资源的投入依然不足，语言智能人才的培养仍旧存在结构性缺陷——了解语言智能算法和应用技术的人对语言资源缺乏认识，对语言规律、语言现象了解不足；从事语言规律研究、参与语言资源建设的人又缺少语言智能科学的素养。这种严重的脱节已经成为影响语言智能发展的瓶颈”。在新文科背景下，学习语言专业的学生要掌握一定的语言智能技术，学习人工智能专业的学生要掌握一定的语言学知识，时代对这一领域复合型人才的需求突显了学习语言智能知识的重要意义。

首先，学习语言专业的学生需要掌握一定的人工智能技术。人工智能已经悄然改变了我们与世界的交流方式。在语言学习和语言研究过程中，人工智能技术可以帮助我们更高效地完成这些任务。比如，AI 可以

通过分析大量的语料库，自动识别和理解语言的模式和规则。这不仅提高了效率，也使得研究更具有客观性和可信度。再如，自然语言处理技术就是 AI 在语言分析中的一个具体应用。语言类学生可以通过这些技术来处理大规模文本数据，以发现趋势、解决问题，省去大量的时间和精力去处理数据。时代发展到今日，以解决问题为导向的跨学科融合已成为科学研究的发展趋势，语言智能作为代表人工智能发展水平的交叉领域，涵盖了计算机科学、人工智能、数据科学和语言学等多个学科领域。学习语言智能可以帮助语言类学生更好地融入跨学科的研究，提升自己的智能素养和水平，跟上科技发展的脚步，为未来做好准备。

另一方面，学习计算机科学的学生或从事人工智能研究的人员也要注重语言学知识的学习。计算机科学和人工智能领域的一个重要分支是自然语言处理（NLP），它涉及计算机理解和生成自然语言的能力。具备语言学知识的人能更好地理解语言结构、语法、语义等概念，从而更有效地进行 NLP 研究。在数据科学和机器学习领域，文本数据的分析和挖掘是常见任务。语言学知识可以帮助研究人员更好地理解文本数据，提取有用的信息，并进行情感分析、主题建模等任务。人工智能研究者可能需要处理多种语言的文本数据。了解语言学原理可以帮助他们更好地理解不同语言之间的差异，从而提高机器翻译和多语言处理系统的性能。知识图谱和语义网技术在人工智能中具有重要地位。语言学知识可以帮助研究人员更好地理解 and 利用知识图谱中的信息。融合计算机科学和语言学的知识可以促进创新性的研究。跨学科合作在解决复杂问题时尤为重要。跨学科的综合知识通常会产生创新性的成果，有助于解决复杂的人工智能挑战。

李宇明（2023）在《语言智能与社会进步》一文中指出，文理分家的教育模式已经不适合大交叉大融合的科学发 展要求。大学本科和研究生教育都要进行深度的文理融合。不仅文科生需要进行必要的理科教育，理工科学生也需要文科教育。特别是人工智能课程，应当成为基础教育的基础课，成为大学各个专业的重要课程。让学生了解人工智能发展趋势，具备人工智能意识，掌握必要的计算科学方法，借助人 工智能处理本专业问题，并有意识地利用专业优势推进人工智能发展。文科和理科之间在学科方法、学科工具与技术训练等方面确实存在很多的差异，但

它们恰恰可以构成问题意识的互补性。语言智能是人工智能皇冠上的明珠，是人工智能之冠，也是新文科建设的新方向。加强语言智能的研究有利于推动计算机科学与技术、认知科学、语言学等多学科的合作，对于创新学科发展体系，培养拔尖创新人才具有重要意义。

近年来，我国高等文科类院校，特别是语言类院校面对快速发展的人工智能技术，高度重视语言智能的研究和人才培养工作，纷纷设立了和语言智能相关的学科和科研机构，在语言机理、语言数据、关键技术、创新应用等方面作了大量卓有成效的工作。在推动语言智能的学科发展、交叉学科人才培养方面取得了积极的进展。然而，很多高校语言类专业未能开设反映语言智能发展现状和关键技术的系列课程，很少向学生系统传授语言智能以及相关学科的知识。尽管一些高校在本科阶段开设了诸如语料库语言学、计算机辅助翻译等课程，这些课程还远不能满足学生智能素养的提升和创新能力的培养。开设诸如语言智能基础、自然语言处理、语言数据处理与应用、语言认知与计算等课程是促进学生的跨学科知识融合、培养具备未来竞争力的高素质人才的时代需求。

## 思考与讨论

1. 为什么说语言智能是人工智能皇冠上的明珠？
2. 如何理解语言认知计算是语言智能的核心要素？
3. 谈一谈语言类专业的学生学习语言智能的重要意义。

## 参考文献

- [1] Gardner, H. *Frames of Mind: Theory of Multiple Intelligence* [M]. New York: Basic Books, 1983.
- [2] Sharp, B. et al. (eds.). 徐金安等译. 自然语言处理的认知方法 [M]. 北京: 机械工业出版社, 2019.
- [3] 艾斌. 语言智能与外语教育——第二届中国外语教育高峰论坛综述 [J]. 中国外语, 2021, 18 (01): 108-111.

- [4] 崔启亮. 人工智能在语言服务企业的应用研究 [J]. 外国语文, 2021, 37 ( 1 ): 26-32+73.
- [5] 邓力, 刘洋. 基于深度学习的自然语言处理 [M]. 北京: 清华大学出版社, 2020.
- [6] 国务院关于印发新一代人工智能发展规划的通知 [D]. 国发〔2017〕35, 2017年7月8日, [http://www.gov.cn/zhengce/content/2017-07/20/content\\_5211996.htm](http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm).
- [7] 胡开宝, 尚文博. 语言学与语言智能 [J]. 华东师范大学学报 ( 哲学社会科学版 ), 2022, 54 ( 02 ): 103-109+176.
- [8] 胡开宝, 田绪军. 语言智能背景下的 MTI 人才培养: 挑战、对策与前景 [J]. 外语界, 2020, ( 2 ): 59-64.
- [9] 胡开宝, 王晓莉. 语言智能视域下外语教育的发展——问题与路径 [J]. 中国外语, 2021, 18 ( 6 ): 4-9.
- [10] 黄河燕, 史树敏等. 人工智能: 语言智能处理 [M]. 北京: 电子工业出版社, 2020.
- [11] 黄立波. 2022. 大数据时代背景下的语言智能与外语教育 [J]. 中国外语, 2020, 19 ( 1 ): 4-9.
- [12] 教育部关于印发《高等学校人工智能创新行动计划》的通知 [D]. 教技〔2018〕3号, 2018年4月2日, [http://www.moe.gov.cn/srcsite/A16/s7062/201804/t20180410\\_332722.html](http://www.moe.gov.cn/srcsite/A16/s7062/201804/t20180410_332722.html).
- [13] 李冠宇, 黄映辉. 智能科学与技术的知识体系: 语义分析的结论 [J]. 计算机教育, 2009, 95 ( 11 ): 61-67.
- [14] 李西, 王霞, 姜孟. 语言智能 赋能未来——第五届中国语言智能大会综述 [J]. 外国语文, 2021, 37 ( 2 ): 141-144.
- [15] 李宇明, 施春宏, 曹文, 王莉宁, 刘晓海, 杨尔弘, 颜伟. “语言资源学理论与学科建设”大家谈 [J]. 语言教学与研究, 2022, ( 2 ): 1-16.
- [16] 李佐文, 梁国杰. 语言智能的学科内容与建设路径 [J]. 外语电化教学, 2022, ( 5 ).
- [17] 李佐文等. 人工智能拓宽话语研究路径 [N]. 光明日报, 2020年7月18日.

- [18] 梁晓波, 邓祯. 美军语言智能处理技术的发展策略与启示 [J]. 国防科技, 2021, 42 (4): 85-91.
- [19] 刘利. 语言智能的学科建设与发展方向 [OL]. [https://www.sohu.com/a/321162236\\_312708](https://www.sohu.com/a/321162236_312708), 2019-06-17.
- [20] 史忠植. 智能科学 (第2版) [M]. 北京: 清华大学出版社, 2013.
- [21] 苏金智. 语言智能水平关乎国家语言能力 [N]. 光明日报, 2021a-08-08 (07版).
- [22] 苏金智. 中国语言智能研发暨语言文化教育传播高峰论坛开幕词 [J]. 汉字文化, 2021b, (1): 4-8.
- [23] 唐杰. 让机器像人一样“思考”: 超越图灵测试的通用机器认知能力 [J]. 智能系统学报, 2020, 15 (6): 1029.
- [24] 唐良元. 2023. 什么是语言智能 [EB/OL]. 千家网. <https://baijiahao.baidu.com/s?id=1774164115319852097#:~:text=%E6%96%87%2F%E5%94%90%E8%89%AF%E5%85%83>.
- [25] 杨尔弘等. 语言智能那些事儿 [A]. 中国语言生活状况报告. 北京: 商务印书馆, 2018, 79-86.
- [26] 语言智能与外语教育协同发展 [J]. 语言教学与研究, 2019 (1): 113.
- [27] 周建设. 语言智能, 在未来教育中扮演什么角色 [J]. 云南教育 (视界综合版), 2019, (4): 45-46.
- [28] 周建设. 加快科技创新 攻关语言智能 [N]. 人民日报, 2020-12-21 (019).
- [29] 周建设. 语言智能研究 [M]. 天津: 天津大学出版社, 2023 (1).
- [30] 周建设等. 语言智能研究渐成热点 [N]. 中国社会科学报, 2017-02-07 (003).
- [31] 周建设, 张文彦. 智能时代的语言学研究 [N]. 中国社会科学报, 2018-09-14 (005).